

1 Convolutional Neural Network for seismic phase
2 classification, performance demonstration over a
3 local seismic network.

4
5 Authors:

6
7 Jack Woollam^{1*}, Andreas Rietbrock^{1,2}, Angel Bueno³, Silvio De Angelis¹.
8

9 ¹University of Liverpool, Liverpool, United Kingdom.

10 ²Karsruhe Institute of Technology, Karlsruhe, Germany.

11 ³University of Granada, Granada, Spain.

12 Abstract

13 Over the last two decades the amount of available seismic data has increased significantly
14 fuelling the need for automatic processing to utilize the vast amount of information contained
15 in such data sets. Detecting seismicity in temporary aftershock networks is one important
16 example, which has become a huge challenge due to the high seismicity rate and dense
17 station coverage. Additionally, the need for highly accurate earthquake locations, to
18 distinguish between different competing physical processes during the post-seismic period,
19 demands even more accurate arrival time estimates of seismic phase. Here we present a
20 Convolutional Neural Network (CNN) for classifying seismic phase onsets for local seismic
21 networks. The CNN is trained on a small dataset for deep-learning purposes (411 events)
22 detected throughout Northern Chile, typical for a temporary aftershock network. In the
23 absence of extensive training data, we demonstrate that a CNN based automatic phase
24 picker can still improve performance in classifying seismic phases, which matches or exceeds
25 that of historic methods. The trained network is tested against an optimised STA/LTA based
26 method ([Rietbrock et al., 2012](#)), in classifying phase onsets for a separate dataset of 3878
27 events throughout the same region. Based on station travel time residuals the CNN out-
28 performs the STA/LTA approach and achieves location residual distribution close to the ones
29 obtained by manual inspection.

30

31 Introduction

32 Accurate detection of earthquake signals generated within the Earth is a fundamental and
33 challenging task in seismology. Traditionally, the optimal method of identifying seismic
34 phases involves a trained analyst manually inspecting seismograms and determining
35 individual phase arrival times. Continuous developments in data acquisition and storage have
36 resulted in vast, unprecedented increases in the volume of available seismic data. For such
37 large-scale datasets, traditional manual picking methods are rendered unfeasible due to the
38 required investment of time and resources; in addition, manual picking incorporates the
39 subjectivity of different analysts which can bias pick accuracy. Further development of reliable
40 automated picking methods are therefore essential to assist seismologists in their efforts to
41 process large-scale datasets.

42 Historic Auto-pickers

43 The pressing need for a reliable automatic phase picker is not new, and numerous methods
44 have been proposed to detect P- and S- wave onsets automatically. The most commonly
45 used method for automatic phase picking is still the STA/LTA approach ([Allen, 1978](#); [Allen, 1982](#); [Earle & Shearer, 1994](#)), which measures the ratio between the energy of the seismic
46 signal over a short-term and a long-term window; any values of the STA/LTA ratio above a
47 defined cut-off threshold represent a phase arrival. [Baer & Kradolfer \(1987\)](#) modified the
48 STA/LTA incorporating an envelope function and a dynamic signal threshold into the
49 characteristic function. There are numerous other approaches, including those based upon
50 higher-order statistics ([Saragiotis et al., 2002, 2004](#); [Küperkock et al., 2010](#)), autoregressive
51 methods ([Leonard & Kennet, 1999](#); [Sleeman & Van Eck, 1999](#); [Rastin et al., 2013](#)), shallow
52 neural networks ([Wang & Teng, 1995](#); [Dai & MacBeth, 1995, 1997](#); [Zhao & Takano, 1999](#);
53 [Gentili & Michelini, 2006](#)), methods which utilise wave polarisation ([Ross & Ben-Zion, 2014](#);
54 [Baillard et al., 2014](#)), and those which utilise pickers in tandem ([Nippres et al., 2010](#)). Whilst

56 there has been extensive development of auto-picker routines, automated picking algorithms
57 cannot currently match the accuracy of an experienced analyst. This is attributed to the
58 complex nature of earthquake source and propagation, with multiple physical processes
59 affecting the wavefield; variations in attenuation, noise-interference, source mechanism and
60 energy-partitioning at interfaces all affect the observed waveform.

61 [Why historic auto-picking routines are typically inferior compared to human analysts](#)
62 Traditional automated picking methods are manually optimized for individual networks and/or
63 even on a station by station basis, fine tuning the ‘characteristic functions’ to distinguish body-
64 wave phases from noise. E.g., triggers can be based on the frequency content of a trace,
65 kurtosis, or some other combination of manually extracted features. One common problem
66 is that that S-wave phases are more difficult to pick as their onset is often masked by the
67 coda of P-waves and manually extracted features will often struggle to identify the S-wave in
68 such instances ([Gomberg et al., 1990](#)).

69 [Advancements in deep-learning](#)

70 Rather than extracting individual features, deep-learning based algorithms focus on learning
71 representations of data, where multiple layers of processing provide varying levels of
72 abstraction ([LeCun et al., 2015](#); [Schmidhuber., 2015](#)). Recent advancements in deep-
73 learning techniques have yielded a suite of procedures which demonstrate ‘super-human’
74 performance when applied to solve problems in fields ranging from computer vision
75 ([Krizhevsky et al., 2012](#)), to speech-recognition ([Hinton et al., 2012](#)). Convolutional Neural
76 Networks (CNNs), are a form supervised machine learning that achieve exceptional results
77 in classifying multi-dimensional inputs such as images, videos, and audio ([Krizhevsky et al.,](#)
78 [2012](#); [Karpathy et al., 2014](#); [LeCun & Bengio, 1995](#)). CNNs apply repeated convolutional and
79 pooling operations to the input data, resulting in a set of learnable filters which automatically
80 ‘engineer’ the appropriate features for classification. The appropriate features are extracted

81 by fine tuning of the network's internal parameters (or weights), via a computer-based
82 optimisation process. The intrinsic properties of CNNs make them an ideal method for natural
83 signal classification ([LeCun et al., 2015](#)). Natural signals often demonstrate local connections
84 between samples, an example being the higher amplitudes observed immediately following
85 an impulsive phase arrival. The major advantage of a CNN approach is how such features
86 are then optimised. Shared weights throughout the network results in the systematic
87 optimisation of decision boundaries to find the best weighted combination of local features to
88 classify phase onsets. Another major factor behind the success of deep-learning methods is
89 that the only required input is a large dataset of labelled examples for training. Within the
90 seismological community, large datasets of labelled data are readily available in the form of
91 manually picked earthquake catalogues for many regions. We are now starting to see the
92 adoption of deep-learning based methods to solve problems in seismological processing (e.g.
93 [Perol et al., 2018](#); [Yoon et al., 2015](#), [Ross et al., 2018](#); [Zhu & Berozza, 2018](#); [Titos et al.,](#)
94 [2018](#)). Preliminary results indicate such methods can match or even surpass human levels
95 of performance in seismic phase classification. So far, CNN approaches have been trained
96 over extensive (~million) catalogues of labelled examples collected over decades ([Ross et](#)
97 [al., 2018](#); [Zhu & Berozza, 2018](#)). We now investigate the dependency of the input data on
98 classification performance by applying a CNN to classify seismic phases, where the network
99 is trained over a relatively small catalogue of events (~11,000 P & S phase pairs). Can a
100 relatively simple CNN architecture display similar performance improvements in the absence
101 of an extensive training dataset? If a feature engineering approach demonstrates
102 generalisation capabilities when trained over a small local dataset with inherent biases, this
103 will further validate the potential of deep-learning based methods over traditional techniques
104 for seismic phase classification.

105 Data

106 The dataset used in training the CNN is a manually picked catalogue of 411 events containing
107 approximately 11,000 P-/S- phase pairs, located throughout the Iquique region of Northern
108 Chile. The training catalogue has also been used to perform a minimum 1D velocity inversion
109 (Woolam et al., 2019). Events occurred between March-May 2014 and are recorded over a
110 network of 65 broadband and short-period stations distributed throughout Northern Chile and
111 Southern Peru; all stations use a sampling frequency of 100Hz (Figure 1).

112 Manual picking of events was performed using Seismic Data eXplorer (SDX) software
113 <http://doree.esc.liv.ac.uk:8080/sdx/>. We process the dataset applying a linear detrend. Whilst
114 the CNN approach is shown to learn the characteristics of P-phases, S-phases, and noise
115 (Zhu & Beroza, 2018), due to our limited training dataset, the CNN network will only be
116 presented with a small portion of noise examples. To limit the potential for the CNN to
117 erroneously identify noise it has not been trained on as phases, and to homogenize the data
118 set due to different instrumentation; we bandpass filter the data between 2-25 Hz, a frequency
119 range which lies in the passband of all instruments deployed.

120 Manual picks are represented probabilistically as a Gaussian function ($\sigma = 1s$, Figure 2),
121 reducing the bias associated with erroneous picks. The σ parameter was determined through
122 manual parameter testing. Larger σ values resulted in the network acting more as an event
123 ‘detector’ where the output probabilities were not impulsive enough to obtain a definitive
124 phase-onset. Values lower than 1 second resulted in a high proportion of ‘miss-picks’ as
125 manual pick errors not captured by the classification vector had a detrimental effect on
126 engineering the appropriate features for phase classification. The dataset is split into training,
127 validation and test batches (with ratios of 80:10:10 respectively).

128 Dataset augmentation and training

129 Deep learning-based classifiers contain a significant number of trainable parameters in the
130 solution space, therefore, an extremely large number of examples are needed to prevent
131 overfitting of the training dataset and to enhance generalisation. Our dataset is relatively
132 small for deep learning purposes. To overcome the limitations associated with a small training
133 dataset we perform several additional processing steps. Events are scaled by multiplication
134 of a value drawn from a lognormal distribution, the ends of the segmented event are tapered
135 to limit impulsive amplitude spikes generated by processing, varying levels of Gaussian noise
136 are then added to each batch¹, resulting in greater variations of signal vs. background noise.
137 The training events are therefore modified to show a range of arrival types, rather than the
138 high-magnitude, well-recorded events that are typically seen in a small catalogue of manually
139 selected earthquakes for further studies. The input window size for the CNN is 6 seconds.
140 To train the CNN, a given input batch is sequentially windowed with a timestep of 0.4
141 seconds. The windows are randomly shuffled before being used in training, preventing the
142 CNN learning any unnecessary temporal order. A small time step is used to increase the total
143 number of events during training; also, having the network learn to recognise the presence
144 of phases at any point in the input window will help the network generalise beyond the training
145 dataset. Formatting the input data in such a way reduces the biases associated with our small
146 dataset and enhances the capability of the network to pick varying types of arrival.

¹ More information on parameters used to aiding generalisation provided in Data Appendix, section A.1

147 Methodology

148 Network architecture

149 The input to the network compromises three one-dimensional windows (x), where each
150 window samples an individual component. For this given input, the network outputs the
151 probability of either P-phase, S-phase, or Noise for each time sample within that window
152 (Figure 4). Probabilities are output by applying the ‘softmax’ or normalised exponential
153 function to the final layer

$$p(Y = i|x) = \frac{e^{\alpha_i(x)}}{\sum_{j=1}^3 e^{\alpha_j(x)}} \quad . \quad (1)$$

154 Where $j = 1,2,3$ represents the P-phase, S-phase and Noise classes, $\alpha(x)$ contains the
155 associated weights for the final layer. The input data are passed through repeated
156 transformations; convolutional operations initially extract the appropriate features to
157 characterise each class, the extracted features then go through repeated re-sampling stages,
158 to output per-class probabilities. At each stage, a Rectified Linear Unit (ReLU) activation
159 function is applied (Nair & Hinton, 2010). The cost function used to train the CNN is given by
160 the negative log-likelihood $NLL(x, \theta)$. For a multi-class classification problem, where each
161 class is characterised as a series of discrete probability distributions, $NLL(x, \theta)$ is also termed
162 the cross-entropy loss function,

$$NLL(x, \theta) = - \sum_{k=1}^3 \sum_{n=0}^{n-N} \log(p(c_k|x_n, \theta)) \quad (2)$$

163 N represents the total number of training instances, c_k corresponds to the class label
164 assigned to the input (x_n), and the network weights (θ). Eq. 2 is minimised using Adaptive
165 Moment Estimation (ADAM, Kingma & Ba, 2014) along with batch training, the network
166 weights are therefore updated at the end of each batch, over n training instances.

Hyperparameter optimisation ([Bergstra & Bengio, 2012](#)) is the derivation of the optimal network parameters and is a major challenge when designing neural network architectures. Parameters such as, number of layers, regularisation of layers, convolutional kernel shape, and the learning rate can all be optimised. Methods to solve this problem consist of, grid search, random search, and manual estimation. As our study aims to demonstrate that a robust CNN can be trained on small datasets, the focus is on efficient implementation over more time-consuming systematic search methods. Once a robust network architecture is derived, a constrained search is performed for the best combination of hyper-parameters.

Our final network architecture consists of 3 convolutional layers, followed by 3 layers of up-sampling (Figure 4). Again, due to the limited nature of the training dataset, the focus for the network architecture is to limit the potential for overfitting. To localise the features corresponding to different classes, convolutional layers apply strided 1D convolutional filters along each component (Figure 3). The stride for the convolutional window is set to 4, this down-samples the time series by a factor of 4 for each layer, reducing the overall number of free parameters and allowing for quicker incorporation of long-term temporal dependencies into the convolution kernel. A dropout parameter is added to the second convolutional layer. Dropout is a regularisation technique which randomly drops weights during training, reducing model complexity ([Srivastava et al., 2014](#)). One-dimensional max-pooling is applied to the final convolutional layer, further reducing the overall number of networks weights.

Picking phases

To obtain P- and S-phase onsets from the CNN output probabilities, we use knowledge of the simple temporal relationships between P- and S-phases to determine onset times (Figure 5).

For the P-phase probability distribution $p = \{p_1, p_2, p_3, \dots, p_n\}$, and the S-phase probability distribution $s = \{s_1, s_2, s_3, \dots, s_n\}$, if P-phase probabilities exceed a defined cut-off threshold

192 p_{cut} the P-phase onset is searched for within the window $[p_{start} \dots p_{end}]$. The P-phase onset
193 is set at the index of the maximum P-phase probability within this range. If the P-phase
194 criterium is met, the corresponding S-phase is searched for within the searched window
195 $[s_{start} \dots s_{end}]$, if $\sum_{i=s_{start}}^{s_{end}} s_i > s_{cut}$ then the S-phase is set at the index of the maximum S-
196 phase probability within search window. Both conditions must be satisfied for an event to be
197 picked, consequently, the ratio of P:S phase picks using these criteria is 1:1. The parameters
198 used in detecting phase onsets are provided in Table 1 – note that all index values are relative
199 to the initial p_{cut} index:

200 Results

201 Predictions

202 The trained network takes a 6 second input window for 3-component data and makes phase
203 predictions for each time-sample within the window. Figure 6 displays a sample of the output
204 phase probabilities for events in the test dataset. The predictions display a clear distinction
205 between P-phases and S-phases, further confirming that deep-learning based classification
206 methods engineer the appropriate features to accurately categorise P, S, and noise classes.
207 This presents a major advantage over historic auto-picking methods which utilise manual
208 feature extraction and often struggled to identify the S-phase. To obtain P/S phase onsets,
209 we apply our autopicker function, with input parameters of Table 1, taking advantage of the
210 simple temporal relationship between P and S phases to assign phase onsets (vertical lines
211 on Figure 6). The phase onsets are then compared against the original manual picks and the
212 residuals are plotted (Figure 7).

213 The residual distribution for the test dataset displays a good agreement in the centre of both
214 the P- and S- residual distribution; however, the CNN has also picked extra events/phases in
215 some waveforms. These extra phase picks may be accurate; however, any additional events
216 are not represented in our classification vectors as a detailed association of individual phases
217 to specific events arriving simultaneously is beyond the scope of this work. This negatively
218 affects the residual distribution and is responsible for several of the large outliers observed.

219 Relocation testing

220 To overcome the issue of extra picked arrival times from simultaneously occurring events
221 biasing our residual comparison, we perform an additional test to remove arrival times from
222 any events overlapping in time. This additional test provides a more consistent assessment
223 of auto-picker performance. We perform an iterative inversion procedure, relocating both the
224 original manual picks and the CNN picks for the initial dataset. The catalogues are relocated

225 using the VELEST routine ([Kissling et al., 1994](#)), which applies a minimum-1D velocity model
226 along with station corrections to solve for hypocentre locations. Hypocentral parameters are
227 solved for all events within the catalogue. When using VELEST, all phase picks within a
228 segmented trace are assigned to a single event during relocation. The large outlier residuals
229 a significant distance (+3s) from the trend are attributed to multiple picked events in the same
230 segmented trace being erroneously classified as a single-event in VELEST. We therefore
231 reject events with RMS residual larger than 3s to remove any picked events overlapping in
232 time. Statistics of the residual distribution for the original manual picks compared against the
233 CNN picks is provided in Table 2.

234 The residual distribution indicates that manually picked P-phases are slightly more accurate
235 than CNN P-phase picks (σ decreased by 0.051s); however, S-phase picks of the CNN
236 approach achieve similar performance to manual picking (σ decreased by 0.019s). We
237 recognize that our training and test data set used for the earthquake location data set are not
238 independent; however, the residual distribution obtained from the CNN methodology is similar
239 to that of the manual picks of an expert seismologist.

240 [Autopicker comparison](#)

241 To further test the CNN picker, we apply the CNN methodology in predicting phase-onsets
242 for a separate catalogue of events throughout Northern Chile, on the same temporary seismic
243 network. Events were initially segmented using an iterative approach based on a STA/LTA
244 trigger ([Rietbrock et al., 2012](#)) and provides a useful test case for the CNN method. The
245 relocation procedure is again applied to compare performance. The initial number of phase-
246 picks for both methods are provided in Table 3.

247 Figure 8 displays an event from the new catalogue picked using the CNN method, multiple
248 event arrivals are again present in the traces. To limit the effect of this issue on our residual

249 comparison, we set both the STA/LTA and CNN method to only pick a single P-/S-phase pair
250 per trace and again use the iterative relocation procedure to assess residual.

251 The relocated hypocentre distribution for both methods are displayed in Figure 9. It can be
252 clearly observed that locations are more clustered in the CNN approach and are better
253 concentrated along the plate interface, indicating the greater consistency in phase picks for
254 the CNN approach. Phase residuals for the relocated events are displayed in Figure 10; we
255 show residuals for both the final catalogues (minimum azimuthal gap $< 220^\circ$) and for only the
256 best-located events (minimum azimuthal gap $< 160^\circ$). Statistics for the residual distributions
257 are displayed in Table 4. Assuming a normal distribution, the CNN method exhibits decreased
258 variance in phase residual for both P- and S- phases when compared to the optimised
259 STA/LTA approach.

260 The relocation residuals (Figure 10) are not just dependent upon accuracy of detected
261 phases, but also on velocity variations not captured in the 1D model or station corrections
262 affecting the residuals. As both catalogues were relocated with the same iterative re-location
263 procedure using the same 1D velocity model and station delay terms, discrepancies in
264 residual distributions should directly reflect the relative consistency of picks in each
265 catalogue. Investigating the residual distribution, the CNN approach has markedly improved
266 both the overall relocation residual (Figure 9), and the variation in residual for both P- and S-
267 phases. In addition to this, the difference in σ for the well-located events is shown to be more
268 accurate for the CNN approach, with σ improving by 0.230s for P-phases and 0.326s for S-
269 phases when compared against the optimised STA/LTA picking approach. The statistics of
270 the residual distribution are also in a similar range to that of the manual picks (see Table 2).

271 Discussion & Future work

272 Accurate and consistent catalogues of phase arrivals are of paramount importance to
273 seismologists, as they typically form the starting point for further seismological investigations.
274 The rapidly increasing amount of seismic data available, along with constant developments
275 in computational capabilities have resulted in the seismological community now increasingly
276 turning to machine-learning based methods to improve the efficiency of seismic processing.
277 As shown, automatic feature engineering approaches such as CNNs hold promise for seismic
278 phase classification, as they only require the 3-component data as an input, and the features
279 engineered from the data combine to detect the general characteristics of P- phases, S-
280 phases and noise. Our experimental results show that even when data are scarce, a simple
281 CNN architecture significantly improves the σ of P- and S-pick residuals, especially for well
282 locatable events (minimum azimuthal gap $< 160^\circ$), resulting in a decrease of 0.230s and
283 0.326s, respectively, when compared against an optimised STA/LTA picking approach
284 ([Rietbrock et al., 2012](#)). The decreased variation in residual, indicates that a CNN based
285 method is more consistent when auto picking, resulting in more accurate hypocentre
286 relocations. We are close to reaching a point where supervised-learning based methods
287 exhibit comparable or even increased performances when compared to manual picking by
288 an expert seismologist ([Zhu & Beroza, 2018](#); [Ross et al., 2018](#)). Until now, supervised
289 learning-based methods have been trained using extensive training datasets (~millions of
290 examples). The results from our work add to the literature of supervised learning-based
291 methods for seismic phase classification and demonstrate that with appropriate
292 considerations regarding overfitting and generalisation, such methods can improve
293 seismological processing workflows, not just for large catalogues, but for varying datasets.
294 Future applications of deep-learning based methods in seismology include deploying such

295 pre-trained systems on poorly monitored areas/areas of interest resulting in improved data
296 recovery, and efficient automation of seismic workflows.

297 Data and Resources

298 All data used in this study can be downloaded from the Incorporated Research Institutes for
299 Seismology (IRIS) data management centre for the temporary network data and also from
300 the GEOFON data repository <https://geofon.gfz-potsdam.de/waveform/archive/>.

301 Acknowledgements

302 This work was partly supported by NERC grant: NE/M005879/1.

303 **References**

- 304 Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces. *Bulletin of*
305 *the Seismological Society of America*, 68(5), pp.1521-1532.
- 306 Allen, R., 1982. Automatic phase pickers: their present use and future prospects. *Bulletin of*
307 *the Seismological Society of America*, 72(6B), pp.S225-S242.
- 308 Baer, M. and Kradolfer, U., 1987. An automatic phase picker for local and teleseismic
309 events. *Bulletin of the Seismological Society of America*, 77(4), pp.1437-1445.
- 310 Baillard, C., Crawford, W.C., Ballu, V., Hibert, C. and Mangeney, A., 2013. An automatic
311 kurtosis-based P-and S-phase picker designed for local seismic networks. *Bulletin of the*
312 *Seismological Society of America*, 104(1), pp.394-409.
- 313 Bergstra, J. and Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal*
314 *of Machine Learning Research*, 13(Feb), pp.281-305.
- 315 Dai, H. and MacBeth, C., 1995. Automatic picking of seismic arrivals in local earthquake data
316 using an artificial neural network. *Geophysical journal international*, 120(3), pp.758-774.
- 317 Dai, H. and MacBeth, C., 1997. The application of back-propagation neural network to
318 automatic picking seismic arrivals from single-component recordings. *Journal of Geophysical*
319 *Research: Solid Earth*, 102(B7), pp.15105-15113.
- 320 Earle, P.S. and Shearer, P.M., 1994. Characterization of global seismograms using an
321 automatic-picking algorithm. *Bulletin of the Seismological Society of America*, 84(2), pp.366-
322 376.
- 323 Gentili, S. and Michelini, A., 2006. Automatic picking of P and S phases using a neural
324 tree. *Journal of Seismology*, 10(1), pp.39-63.

325 Gombert, J.S., Shedlock, K.M. and Roecker, S.W., 1990. The effect of S-wave arrival times
326 on the accuracy of hypocenter estimation. *Bulletin of the Seismological Society of*
327 *America*, 80(6A), pp.1605-1628.

328 Hayes, G.P., Wald, D.J. and Johnson, R.L., 2012. Slab1. 0. A three-dimensional model of
329 global subduction zone geometries. *Journal of Geophysical Research: Solid Earth*, 117(B1).

330 Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L., 2014. Large-
331 scale video classification with convolutional neural networks. In *Proceedings of the IEEE*
332 *conference on Computer Vision and Pattern Recognition* (pp. 1725-1732).

333 Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint*
334 *arXiv:1412.6980*.

335 Kissling, E., Ellsworth, W.L., Eberhart-Phillips, D. and Kradolfer, U., 1994. Initial reference
336 models in local earthquake tomography. *Journal of Geophysical Research: Solid*
337 *Earth*, 99(B10), pp.19635-19646.

338 Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep
339 convolutional neural networks. In *Advances in neural information processing systems* (pp.
340 1097-1105).

341 Küperkoch, L., Meier, T., Lee, J., Friederich, W. and EGELADOS Working Group, 2010.
342 Automated determination of P-phase arrival times at regional and local distances using higher
343 order statistics. *Geophysical Journal International*, 181(2), pp.1159-1170.

344 LeCun, Y. and Bengio, Y., 1995. Convolutional networks for images, speech, and time
345 series. *The handbook of brain theory and neural networks*, 3361(10), p.1995.

346 LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, 521(7553), p.436.

Leonard, M. and Kennett, B.L.N., 1999. Multi-component autoregressive techniques for the analysis of seismograms. *Physics of the Earth and Planetary Interiors*, 113(1-4), pp.247-263.

Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-814).

Nippres, S.E.J., Rietbrock, A. and Heath, A.E., 2010. Optimized automatic pickers: application to the ANCORP data set. *Geophysical Journal International*, 181(2), pp.911-925.

Perol, T., Gharbi, M. and Denolle, M., 2018. Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2), p.e1700578.

Rastin, S.J., Unsworth, C.P., Benites, R. and Gledhill, K.R., 2013. Using real and synthetic waveforms of the Matata swarm to assess the performance of New Zealand GeoNet phase pickers. *Bulletin of the Seismological Society of America*, 103(4), pp.2173-2187.

Rietbrock, A., Ryder, I., Hayes, G., Haberland, C., Comte, D., Roecker, S. and Lyon-Caen, H., 2012. Aftershock seismicity of the 2010 Maule Mw= 8.8, Chile, earthquake: Correlation between co-seismic slip models and aftershock distribution?. *Geophysical Research Letters*, 39(8).

Ross, Z.E. and Ben-Zion, Y., 2014. Automatic picking of direct P, S seismic phases and fault zone head waves. *Geophysical Journal International*, 199(1), pp.368-381.

Ross, Z.E., Meier, M.A. and Hauksson, E., 2018. P-wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*.

367 Saragiotis, C.D., Hadjileontiadis, L.J. and Panas, S.M., 2002. PAI-S/K: A robust automatic
 368 seismic P phase arrival identification scheme. *IEEE Transactions on Geoscience and*
 369 *Remote Sensing*, 40(6), pp.1395-1404.

370 Saragiotis, C.D., Hadjileontiadis, L.J., Rekanos, I.T. and Panas, S.M., 2004. Automatic P
 371 phase picking using maximum kurtosis and/spl kappa/-statistics criteria. *IEEE Geoscience*
 372 *and Remote Sensing Letters*, 1(3), pp.147-151.

373 Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural networks*, 61,
 374 pp.85-117.

375 Sleeman, R. and Van Eck, T., 1999. Robust automatic P-phase picking: an on-line
 376 implementation in the analysis of broadband seismogram recordings. *Physics of the earth*
 377 *and planetary interiors*, 113(1-4), pp.265-275.

378 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014.
 379 Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine*
 380 *Learning Research*, 15(1), pp.1929-1958.

381 Titos, M., Bueno, A., García, L. and Benítez, C., 2018. A Deep Neural Networks Approach to
 382 Automatic Recognition Systems for Volcano-Seismic Events. *IEEE Journal of Selected*
 383 *Topics in Applied Earth Observations and Remote Sensing*, 11(5), pp.1533-1544.

384 Wang, J. and Teng, T.L., 1995. Artificial neural network-based seismic detector. *Bulletin of*
 385 *the Seismological Society of America*, 85(1), pp.308-319.

386 Woollam, J., Rietbrock, A., et al., Determination of a minimum 1D seismic velocity model for
 387 the Iquique region of Northern Chile. *Geophysical Journal International*, submitted.

388 Yoon, C.E., O'Reilly, O., Bergen, K.J. and Beroza, G.C., 2015. Earthquake detection through
389 computationally efficient similarity search. *Science advances*, 1(11), p.e1501057.

390 Zhao, Y. and Takano, K., 1999. An artificial neural network approach for broadband seismic
391 phase picking. *Bulletin of the Seismological Society of America*, 89(3), pp.670-680.

392 Zhu, W. and Beroza, G.C., 2018. PhaseNet: A Deep-Neural-Network-Based Seismic Arrival
393 Time Picking Method. *arXiv preprint arXiv:1803.03211*.

394

395 Appendix

396 A.1 Hyper-parameters and data generalisation parameters.

397 Lognormal distribution used to scale individual event amplitudes is given by

398

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)}, \quad x > 0. \quad (\text{A.1})$$

399 We set $\mu = 0$, and $\sigma = 0.25$, and sample the output probability distribution of eq. A.1, each
400 sample is then used as a scale factor for event amplitudes.

401 Mailing addresses
402 Jack Woollam
403 Jack.Woollam@liverpool.ac.uk
404
405 Andreas Rietbrock
406 andreas.rietbrock@kit.edu
407
408 Angel Bueno
409 angelbueno@ugr.es
410
411 Silvio De Angelis
412 S.De-Angelis@liverpool.ac.uk
413
414

Figure captions

Figure 1 | Distribution of manually picked events throughout Northern Chile, stations are indicated by white triangles, event hypocentres are plotted as a function of depth.

Figure 2 | An example of input data (top) and classification data (bottom), inputs to the CNN are 3-component traces, linear-detrended, bandpass filtered between 2-25Hz. The associated classification vector for P-pick and S-pick are represented probabilistically as a Gaussian with $\sigma = 1$ s.

Figure 3 | Schematic displaying how strided 1D convolutions quickly incorporate the long-term temporal dependencies of the input data into the convolution kernel.

Figure 4 | Overall CNN architecture, displaying the sequential convolution and re-sampling operations applied to the input window.

Figure 5 | Displaying how the temporal relationship between P- and S-phases is used to identify phase onsets from the output CNN probabilities. Solid lines correspond to the output P-/S-phase probabilities; vertical dashed lines indicate phase onsets and the phase-type is labelled above each vertical dashed line. Vertical dotted lines indicate the start or end of a P-/S-phase search window, where the corresponding labels are again presented at the top of each line. The horizontal dotted line represents the p_{cut} parameter used in determining phase onsets.

Figure 6 | Output CNN prediction probabilities when applied to identify phase onsets for the test dataset, phase onsets are indicated by vertical lines.

438 **Figure 7** | Residual of CNN predicted phase onsets vs. original manual picks for the test dataset.

439

440 **Figure 8** | Demonstrating the CNN auto picker performance on a new dataset for Northern Chile,
441 where events were segmented using an STA/LTA trigger (Rietbrock et al., 2012). We only allow the
442 auto picker to pick the presence of a single P-/S-phase per trace, to prevent relocation errors.

443

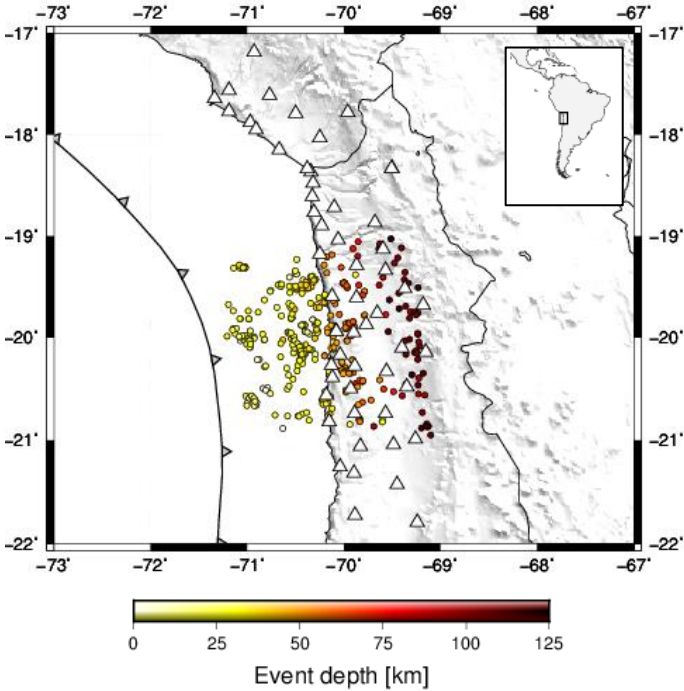
444 **Figure 9** | Hypocentre relocation comparison for the STA/LTA auto picked catalogue (top) against the
445 CNN auto picked catalogue (bottom) event relocations are plotted as a function of RMS residual, slab
446 profile is provided by Hayes et al., (2012).

447

448 **Figure 10** | Both auto picking methods phase residuals following hypocentral relocations, plotted for
449 well-located events (minimum azimuthal gap $< 160^\circ$) and the for entire relocated catalogues
450 (minimum azimuthal gap $< 220^\circ$).

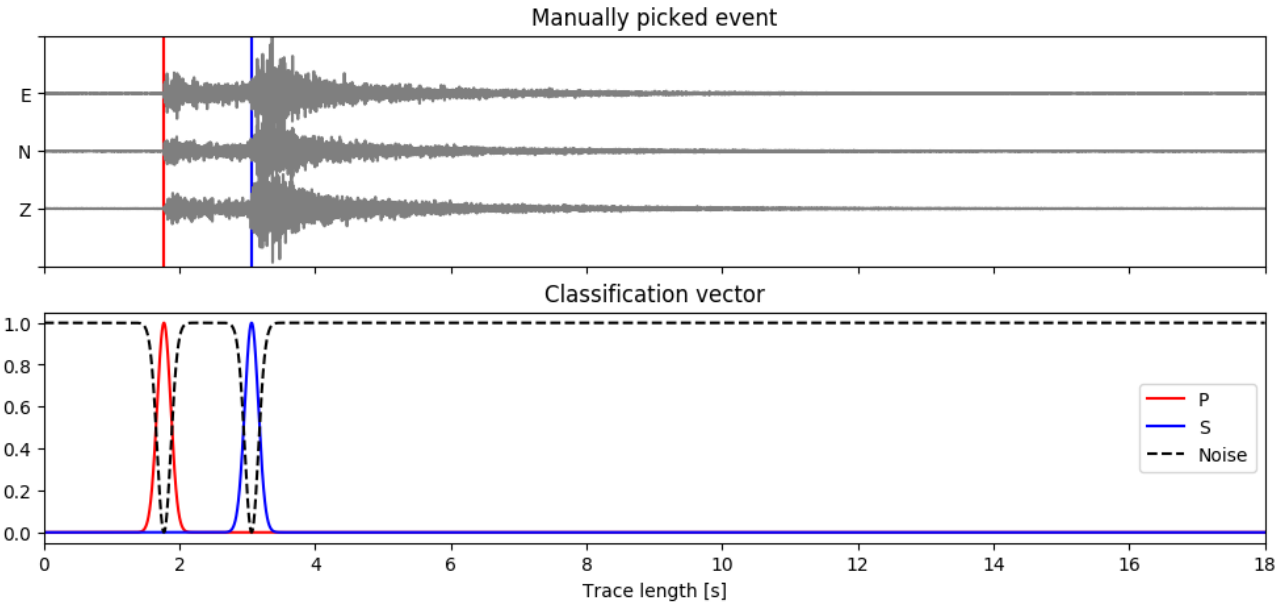
451 Figures

452 Figure 1



453

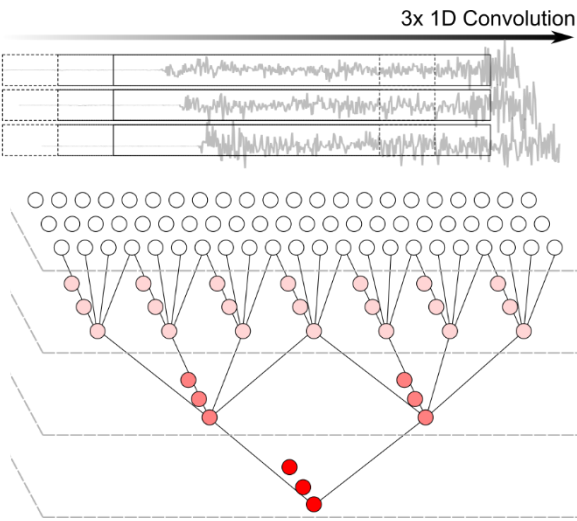
454 Figure 2



455

456

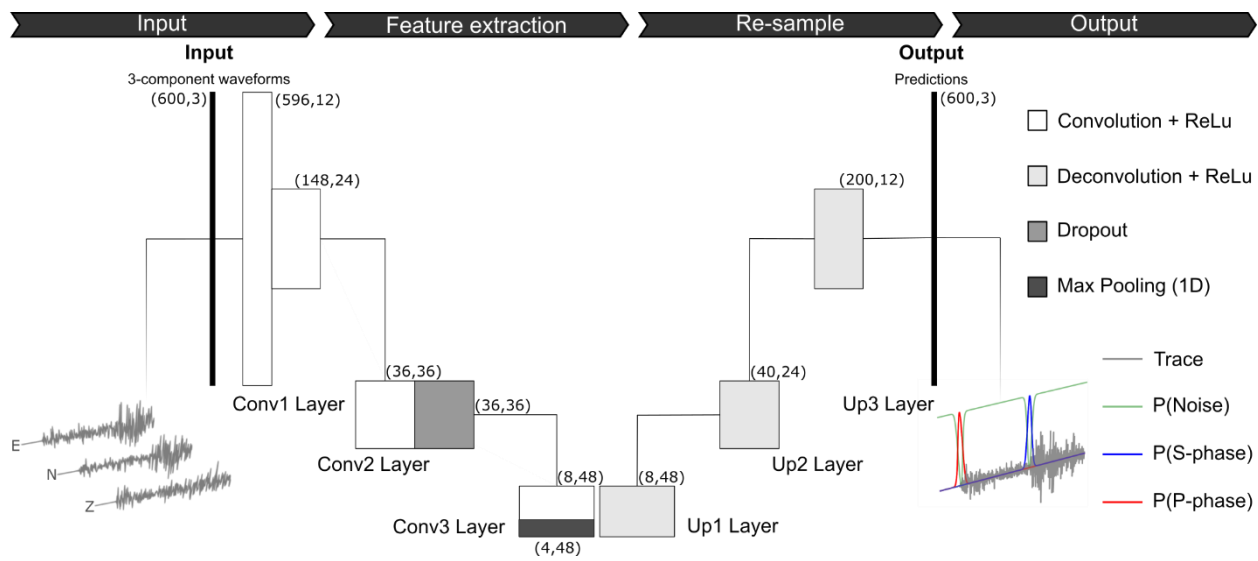
457 Figure 3



458

459

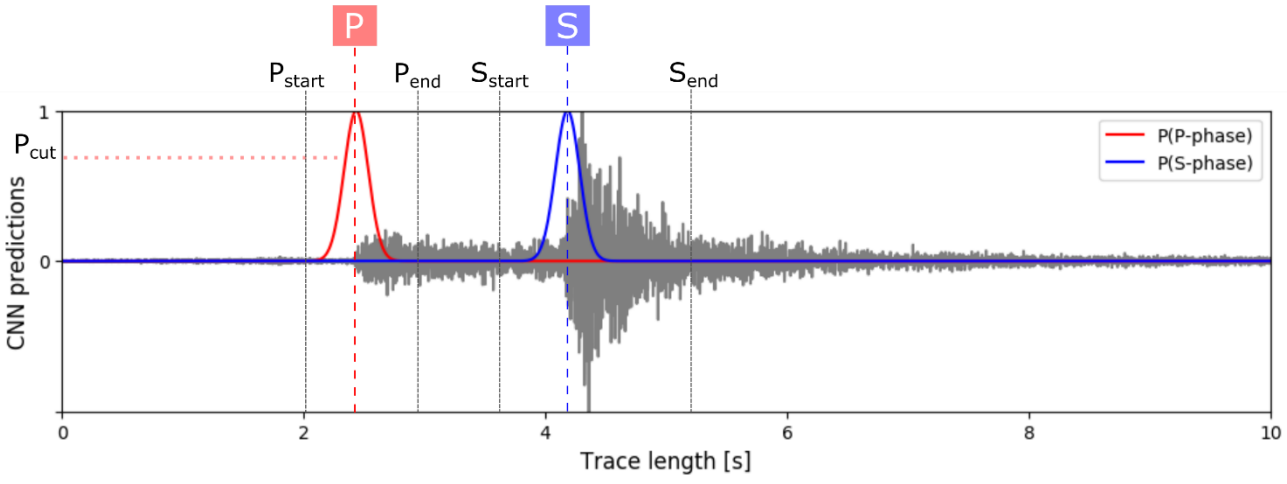
460 Figure 4



461

462

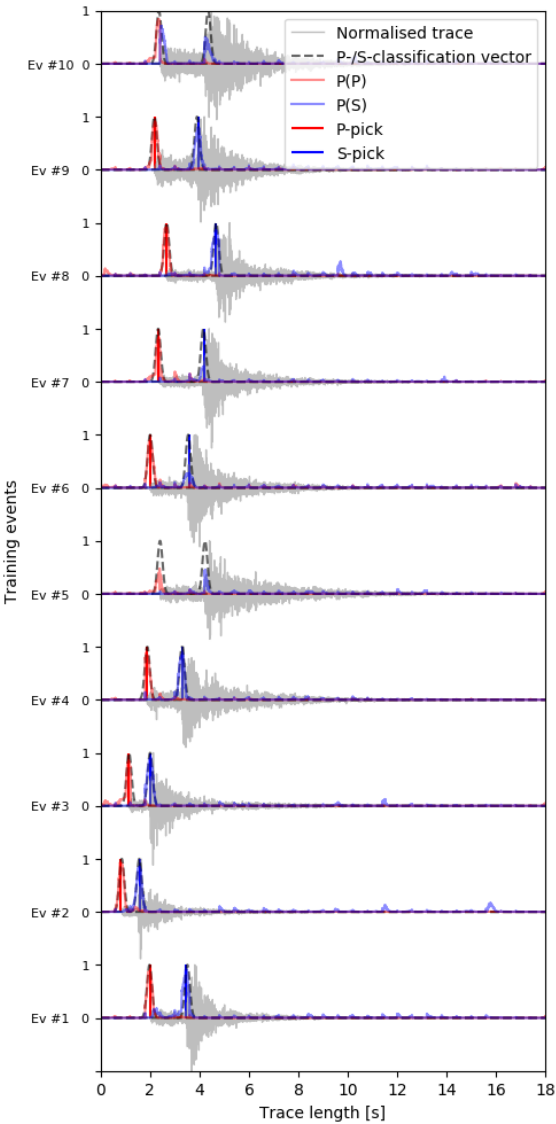
463 Figure 5



464

465

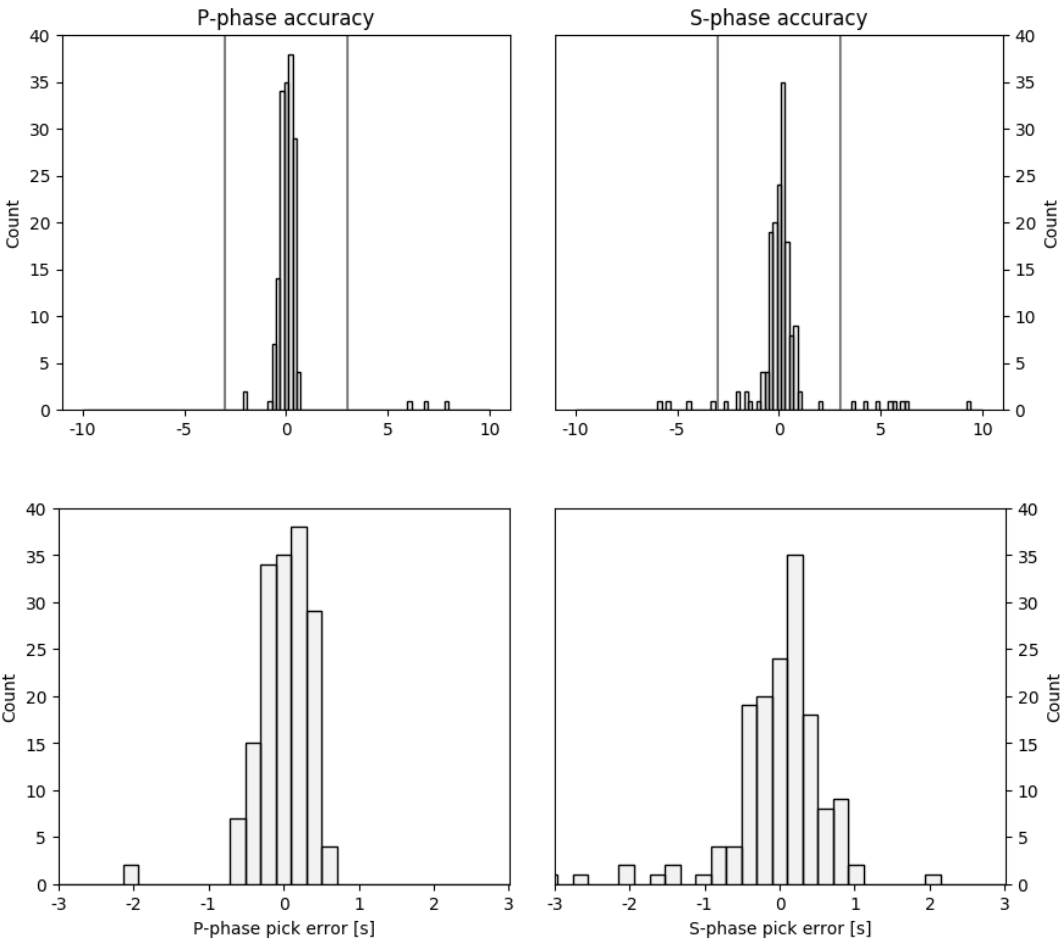
466 Figure 6



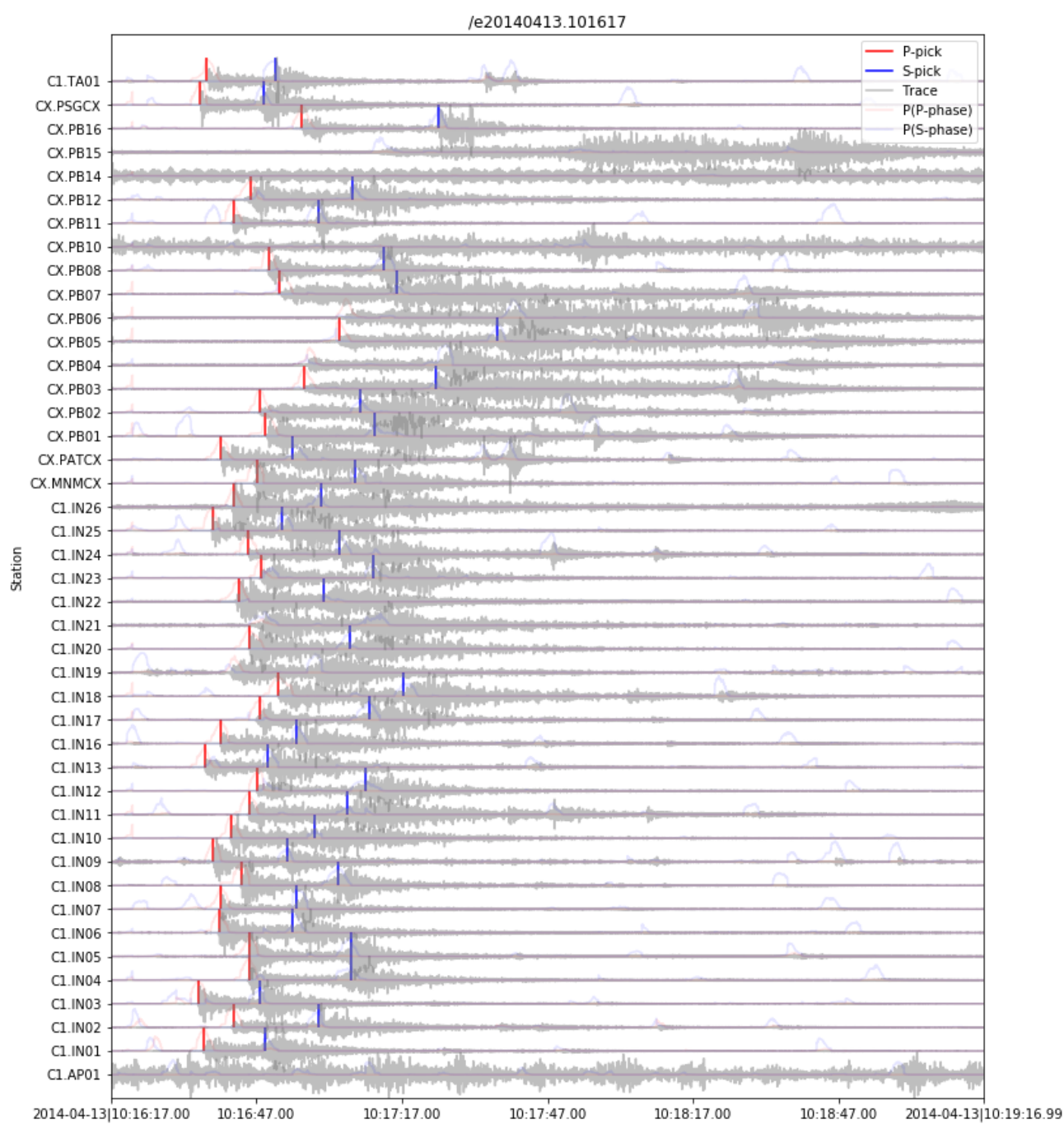
467

468

469 Figure 7



473 Figure 8

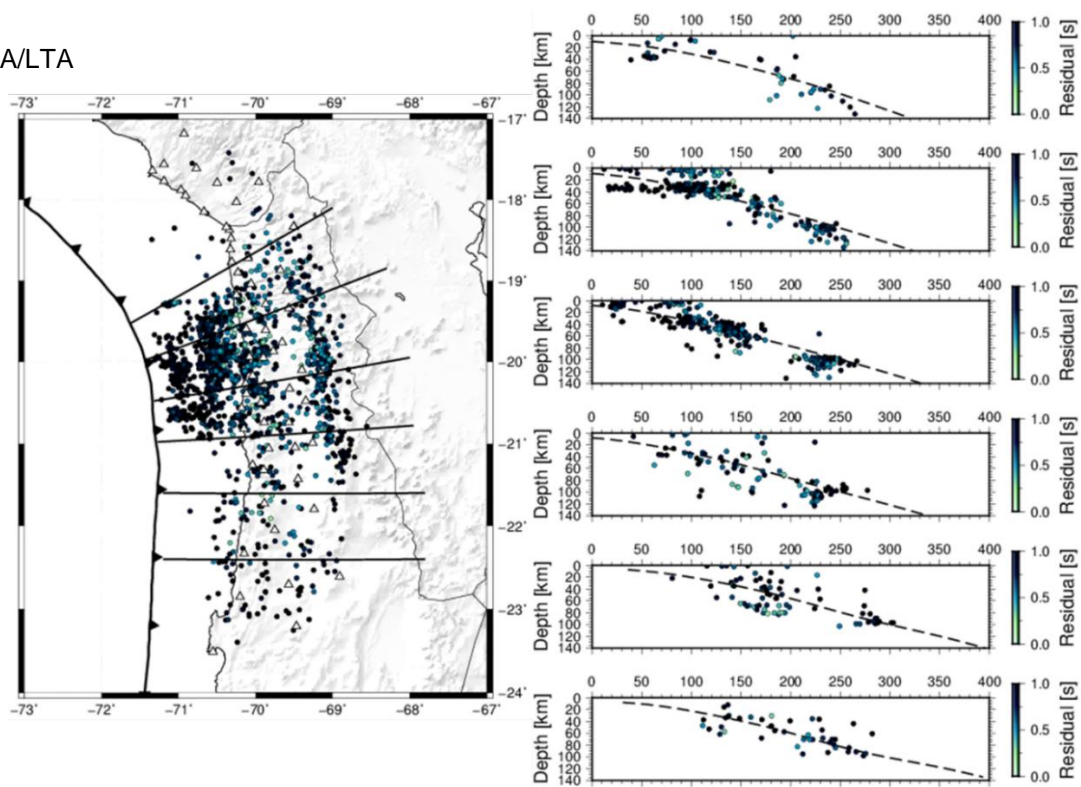


474

475

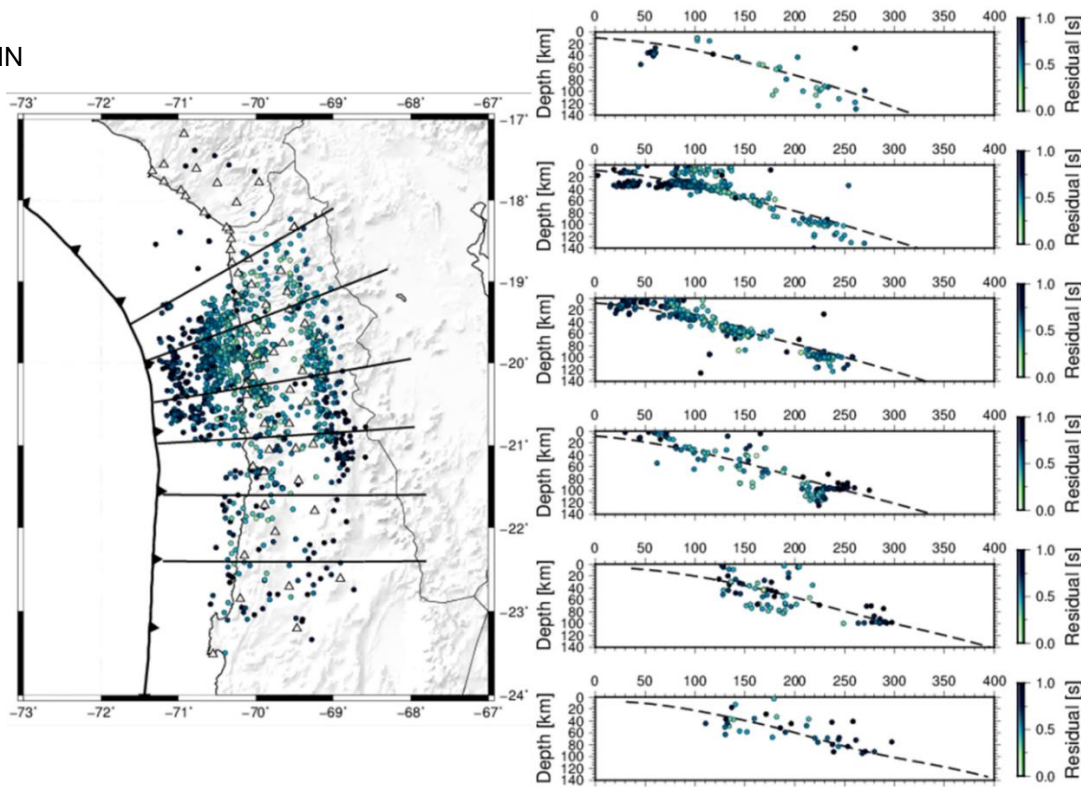
476 Figure 9

STALTA



477

CNN



478

479

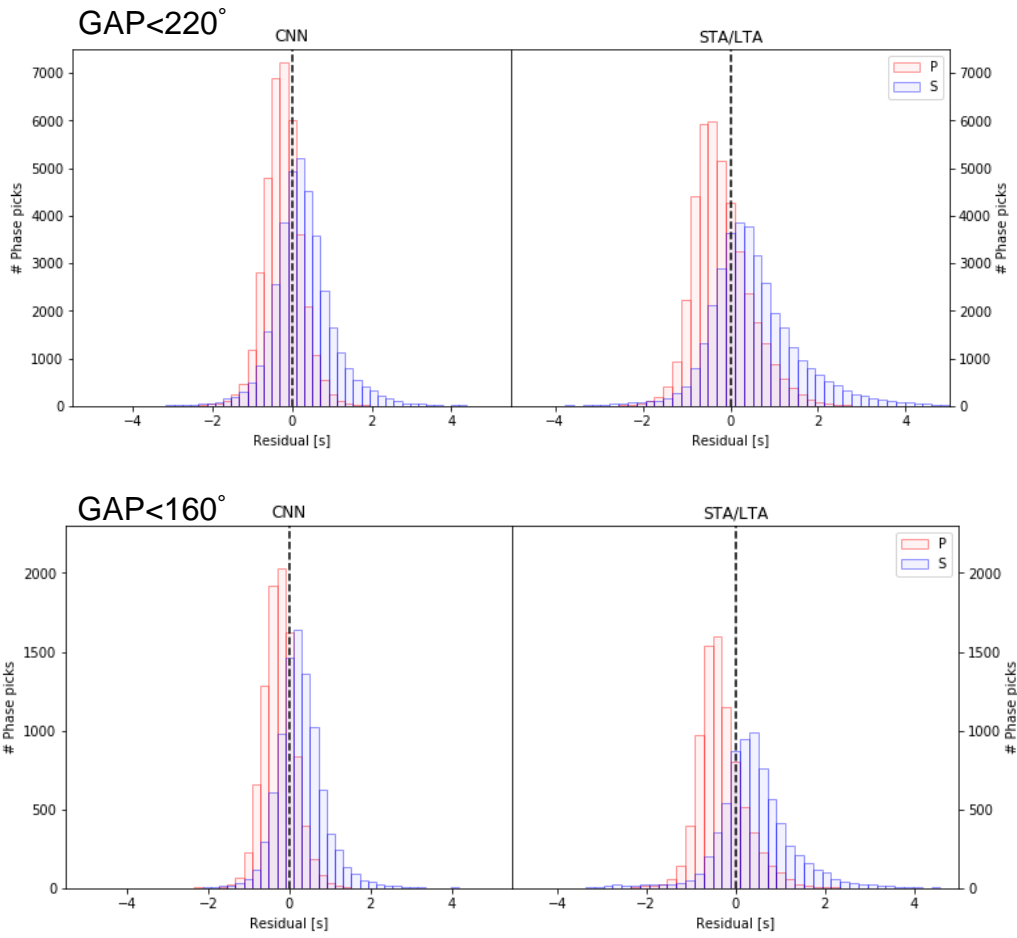
480 Figure 10

481

482

483

484



485 **Tables**

486 **Table 1**

487 **Table 1** | Parameters applied to the autopicker function, which takes advantage of the temporal
488 relationship between phases to identify phase onsets, all start/end indexes are given in samples
489 (where sampling rate for all instruments = 100Hz).

p_{cut}	0.75	s_{sum}	5
p_{start}	-200	s_{start}	500
p_{end}	200	s_{end}	4000

490 **Table 2**

491 **Table 2** | Statistics of residual distribution for original manually picked catalogue, both for the original
492 manual picks and the CNN methodology picks.

GAP < 220°		CNN	Manual
P	μ	-0.261	-0.124
	σ	0.445	0.394
S	μ	0.282	0.390
	σ	0.749	0.730

493 **Table 3**

494 **Table 3** | Overall auto picks on a separate catalogue of new events throughout Northern Chile.

	STA/LTA	CNN
P	72,655	77,623
S	63,353	77,623
Total	136,008	155,246

495

496 Table 4

497 Table 4 | Statistics of residual distribution for GAP < 220° and GAP < 160°.

		GAP < 220°		GAP < 160°	
		CNN	STA/LTA	CNN	STA/LTA
P	μ	-0.238	-0.216	-0.247	-0.333
	σ	0.487	0.696	0.393	0.623
S	μ	0.277	0.539	0.270	0.435
	σ	0.780	1.081	0.596	0.922

498

499

500